Mammalian Genome 8, 287–289 (1997).

Mammalian Genome

Molecular cloning and mapping of a novel developmentally regulated human C2H2-type zinc finger

Kevin G. Becker,[1] Rachel D. Canning,[1] James W. Nagle,[2] Anindya M. Dehejia,[3] Mihael H. Polymeropoulos,[3] Insong J. Lee,[4] Ameer M. Gado,[1] William E. Biddison,[1] Paul D. Drew[1]

[1]Neuroimmunology Branch, Building 10, Room 5B16, National Institute of Neurological Diseases and Stroke, National Institutes of Health, Bethesda, Maryland 20892, USA
[2]Laboratory of Neurogenetics, National Institute of Neurological Diseases and Stroke, National Institutes of Health, Bethesda, Maryland 20892, USA
[3]Gene Mapping Unit, Laboratory of Genetic Disease Research, National Center for Human Genome Research, National Institute of Health, Bethesda, Maryland 20892, USA
[4]Laboratory of Neurogenetics, National Institute of Alcohol Abuse and Alcoholism, National Institutes of Health, Bethesda, Maryland 20892, USA

Received: 3 September 1996 / Accepted: 25 November 1996

Zinc finger domains are found in a variety of protein families including steroid receptor (Beato 1989), ring finger (Freemont 1993), and C2H2-type (Pieler and Bellefroid 1994) zinc finger proteins. C2H2-type zinc finger proteins possess two conserved cysteines that are part of an antiparallel beta sheet and two conserved histidines which are part of an alpha-helix, coordinated by a central zinc atom to form a globular domain. C2H2 zinc finger proteins comprise one of the largest families of proteins known,

Correspondence to: P.D. Drew

```
   1        GAGGCACCTTTCAGAAGTTATGTGGACACTGCCTCGTTTACACAGAGTTGCATA
  54  GTCCATGTGTCGGAGAAACCCTTTACCTGCAGGGAGATCAGGAAAGACTTCCTGGCCAAC
 114  ATGAGGTTTCTCCATCAAGACGCCACTCAAACAGGGGAGAAGCCAAATAACAGTAACAAG
      M  R  F  L  H  Q  D  A  T  Q  T  G  E  K  P  N  N  S  N  K    20
 174  TGTGCGGTGGCCTTTTACAGTGGAAAAAGTCATCACAACTGGGGAAAATGCAGTAAAGCC
      C  A  V  A  F  Y  S  G  K  S  H  H  N  W  G  K  C  S  K  A    40
 234  TTTAGCCACATAGACACACTTGTTCAGGACCAGAGAATCCTCACTAGAGAAGGACTTTTT
      F  S  H  I  D  T  L  V  Q  D  Q  R  I  L  T  R  E  G  L  F    60
 294  GAGTGCAGTAAATGTGGGAAAGCATGTACGCGAAGATGTAACCTCATTCAGCACCAGAAA
      E  C  S  K  C  G  K  A  C  T  R  R  C  N  L  I  Q  H  Q  K    80
 354  GTCCACAGTGAAGAAAGGCCTTATGAATGCAATGAATGTGGAAAATTCTTTACCTACTAC
      V  H  S  E  E  R  P  Y  E  C  N  E  C  G  K  F  F  T  Y  Y   100
 414  TCCAGTTTCATTATACATCAGAGAGTTCATACTGGAGAAAGGCCTTATGCGTGCCCTGAA
      S  S  F  I  I  H  Q  R  V  H  T  G  E  R  P  Y  A  C  P  E   120
 474  TGTGGGAAATCGTTTAGTCAGATATACAGCCTCAATAGCCATAGGAAAGTTCACACTGGA
      C  G  K  S  F  S  Q  I  Y  S  L  N  S  H  R  K  V  H  T  G   140
 534  GAAAGGCCTTATGAATGTGGGGAATGTGGGAAATCTTTTAGCCAAAGGTCCAACCTCATG
      E  R  P  Y  E  C  G  E  C  G  K  S  F  S  Q  R  S  N  L  M   160
 594  CAGCATCGCAGAGTTCACACTGGAGAAAGGCCTTATGAATGCAGCGAATGTGGGAAATCT
      Q  H  R  R  V  H  T  G  E  R  P  Y  E  C  S  E  C  G  K  S   180
 654  TTTAGCCAAAACTTTAGCCTGATCTACCACCAGAGAGTTCACACTGGAGAAAGACCTCAT
      F  S  Q  N  F  S  L  I  Y  H  Q  R  V  H  T  G  E  R  P  H   200
 714  GAGTGCAATGAATGTGGAAAATCCTTTAGCCGAAGCTCCAGCCTCATTCACCACCGGAGA
      E  C  N  E  C  G  K  S  F  S  R  S  S  S  L  I  H  H  R  R   220
 774  CTTCACACTGGAGAAAGACCCTATGAGTGCAGTAAATGTGGGAAGTCATTTAAGCAAAGC
      L  H  T  G  E  R  P  Y  E  C  S  K  C  G  K  S  F  K  Q  S   240
 834  TCCAGCTTCAGTTCACATCGGAAAGTCCACACAGGGGAAAGGCCTTATGTGTGTGGGGAA
      S  S  F  S  S  H  R  K  V  H  T  G  E  R  P  Y  V  C  G  E   260
 894  TGTGGGAAATCCTTTAGCCATAGCTCCAACCTTAAGAACCACCAGAGAGTTCACACTGGA
      C  G  K  S  F  S  H  S  S  N  L  K  N  H  Q  R  V  H  T  G   280
 954  GAAAGACCTGTTGAGTGCAGTGAATGTAGCAAATCCTTTAGCTGTAAATCTAACCTCATT
      E  R  P  V  E  C  S  E  C  S  K  S  F  S  C  K  S  N  L  I   300
1014  AAACACCTGAGAGTTCACACTGGAGAAAGGCCTTATGAGTGCAGTGAATGTGGGAAATCC
      K  H  L  R  V  H  T  G  E  R  P  Y  E  C  S  E  C  G  K  S   320
1074  TTTAGCCAAAGTTCTAGCCTCATTCAACACCGCAGAGTTCACACGGGAAAAAGGCCTTAT
      F  S  Q  S  S  S  L  I  Q  H  R  R  V  H  T  G  K  R  P  Y   340
1134  CAGTGCAGTCAATGTGGGAAATCCTTTGGCTGCAAATCTGTCCTCATTCAACACCAGAGA
      Q  C  S  Q  C  G  K  S  F  G  C  K  S  V  L  I  Q  H  Q  R   360
1194  GTTCACATTGGAGAAAAGCCTTAGCTGTACTGAGAATATGCAATTTCCTTTTAGTGTAAT
      V  H  I  G  E  K  P  *                                       380
1254  TATACTGAAGGAGTACACCTGTGAGAGAGACAAGTACCTGATTTGGAAGCCCCAACATCT
1314  AAGGATATACAGTGGGCGGATTCCCCTTAAGTTCCAGGTATGTGTTACACTTTCTAACAT
1374  GCCATTTAGAAAGTGTTAGACTTTCTCACCTGCCATTTATGGCTCTTGCCGTTTATGTCA
1434  CTGACAGTTTCTGAGGCAGAAGCCGTATCATGTCTACCACCTGTGAGGTCCACACAGTGT
1494  GTATCATTTACCTCCTGAACCTGCTCAAGGAAGCAGACCTCTGCTTCTCCCCATTTGCTA
1554  GAAGAAATCATGAATAGTCTGAGTCTTCCTCTCTGACAAGTTAGGGCATGGACTTGACCC
1614  AGCTTTGTGCCAGAGAACCCAATATGAGTGTTGTTGGCAGCTTGCCAAGAAGGACTGTCT
1674  TTTTCAAGACATACTGGTTTCATGTGACACCTCCATGGATTTTTTTCCAGCCTCTAAGTC
1734  ACCAACTTGGGAACTGCTTGTCTCACGTTGCTTTGTTTTTTACAATAATAAAAGCATTAT
1794  TATTTAAGCTAAAAAACCATTTCACTATCATATTCATCGGCGTAAATCTAACTTTCTTCC
1854  CACAACACTTTCTCGGCCTATCCGGAATGCCCCGACGTTACTCGGACTACCCCGATGCAT
1914  ACACCACATGAAACATCCTATCATCTGTAGGCTCATTCATTTCTCTAACAGCAGTAATAT
1974  TAATAATTTTCATGATTTGAGAAGCCTTCGCTTCGAAGCGAAAAGTCCTAATAGTAGAAG
2034  AACCCTCCATAAACCTGGAGTGACTATATGGATGCCCCCCACCCTACCACACATTCGAAG
2094  AACCCGTATACATAAAAT  2111
```

A

```
C2H2-25  CSKCGKACTRRCNLIQHQKVHSEERPYECNECGKFFTYYSSFIIHQRVHTGERPYACPECGKSFSQIYSLNSHRKVH
ZNF 134  -L---R-FSQSS-FLR-----TQV-----SQ---S-SRS-ALIQ-W-----E---E-SE--RA-NNNSN-AQ-Q---
ZNF 132  --E---TFS-KD--T--KRI-TG-M--K------Y-SHH-NL-V-----N-A---K-SD---V-RHKST-VQ-ESI-

C2H2-25  TGERPYECGECGKSFSQRSNLMQHRRVHTGERPYECSECGKSFSQNFSLIYHQRVHTGERPHECNECGKSFSRSSSLIHHRRLH
ZNF 134  -----F--S---RD---S-H-LR-QK-------F--CD---A--NSST--Q--K----Q--Y--SE-R-S--------Q-W-I-
ZNF 132  ---N--D-SD-----GHKYT-IK-Q-I--ESK-F--I----F--RSSDY-A----------FV-SK---D-I-T-H-VR-Q-V-

C2H2-25  TGERPYECSKCGKSFKQSSSFSSHRKVHTGERPYVCGECGKSFSHSSNLKNHQRVHTGERPVECSECSKSFSCKSNLIKHLRVH
ZNF 134  ---K-----E---A-AH--TLIE-WR---K----E-N----F--QN-I-IK--K-----K-YK----G-F--R--S--C-W---
ZNF 132  ---------E---AYSL--HLNR-Q----AG-L*

C2H2-25  TGERPYECSECGKSFSQSSSLIQHRRVHTGKRPYQCSQCGKSFGCKSVLIQHQRVHIGEKP*
ZNF 132  ------------RA--SN-H-VR-Q----QE---E-I----A-SER-T-VR--K--TR-RTYECSQCGKLFSHLCNLAQHKKIHT*
```

B

Fig. 1. **(A)** Nucleotide and deduced amino acid sequence of C2H2-25. C2H2-25 cDNA was isolated from a human hippocampus library (Stratagene #936205, La Jolla, Calif.) with a degenerate oligodeoxyribonucleotide probe made against a conserved domain common to C2H2-type zinc finger proteins (Becker et al. 1995). The clone was sequenced in both directions following creation of nested deletions on an Applied Biosystems 373A automated sequencer. The nucleotide and amino acid sequence was compiled with GCG software and compared with the non-redundant database at NCBI, National Library of Medicine, by use of the Blast algorithm. The nucleotide (left) and amino acid (right) sequences are numbered in the margins. The C2H2-type zinc finger domains are underlined. The sequence appears in Genbank (accession # U38904). **(B)** Amino acid comparison of C2H2-25, ZNF-132, and ZNF-134. Amino acid identity is denoted by a dash.

with an estimated 300–500 genes in the human genome (Bellefroid et al. 1989). The protein family is typified by the RNA polymerase III transcription factor TFIIIA (Ginsberg et al. 1984; Brown et al. 1985; Miller et al. 1985) and the gap gene product Kruppel (Schuh et al. 1986).

The function of most C2H2-type zinc finger proteins is unknown. However, individual proteins have been shown to be important in development, tumorigenesis, RNA metabolism, and chromatin assembly. Some C2H2 zinc finger proteins are believed to affect gene expression through sequence-specific binding to DNA and/or RNA and through protein-protein interactions (Lee et al. 1993). In addition to the conserved zinc finger domain, some C2H2-type zinc finger proteins contain the conserved amino acid sequence TGEKP between adjacent fingers (Schuh et al. 1986), as well as KRAB (Bellefroid et al. 1991), POZ/tramtrack (Harrison and Travers 1990), and homeodomains (Fortini et al. 1991) found outside of the zinc finger domain. Recently, we reported the isolation of 118 novel C2H2-type zinc finger encoding cDNAs (Becker et al. 1995). Here, we describe the mapping and molecular characterization of one of these clones, C2H2-25.

C2H2-25 cDNA was isolated from a human hippocampus cDNA library (Stratagene #936205). The clone was sequenced in both directions with an Applied Biosystems 373A automated sequencer following the creation of nested plasmid deletions. The C2H2-25 cDNA is 2111 nucleotides long and includes an open reading frame of 380 amino acids with a calculated molecular weight of approximately 44 kDa. C2H2-25 contains 11 zinc finger domains of the C2H2-type, approximately evenly distributed throughout the protein coding region of the clone. C2H2-25 cDNA contains a 5′ untranslated sequence and an initiator methionine, and probably represents the full-length clone, as the mRNA size on Northern analysis is approximately 2 kb (Fig. 1A). This methionine is in consensus for a eukaryotic translational start codon, with a purine (A) at −3 and a G at −6, relative to the ATG.

C2H2-25 is most related by sequence homology to two other zinc finger genes, ZNF-132 and ZNF-134. Amino acid comparisons of the zinc finger regions of these three proteins are shown in Fig. 1B. ZNF-132 and ZNF-134 map to Chromosome (Chr) 19q13.4 (Tommerup and Vissing 1995) and with C2H2-25 may represent a cluster of related zinc finger genes.

Southern blot analysis was performed in order to determine the evolutionary conservation of the gene encoding C2H2-25. This analysis indicates that the C2H2-25 gene is more highly conserved in primates and is also present in other vertebrate species (Fig. 2). C2H2-25 gene is not present in drosophila or yeast. Relatively low hybridization stringency was used in Southern analyses in order to detect the C2H2-25 gene in non-primate vertebrates. At this hybridization stringency, C2H2-25, and probably homologous genes, are detected in primates. Southern analyses at higher hybridization stringency with a probe prepared against the 3′ UTR of C2H2-25 detected a single band in human Southern blots. This supports the idea that the

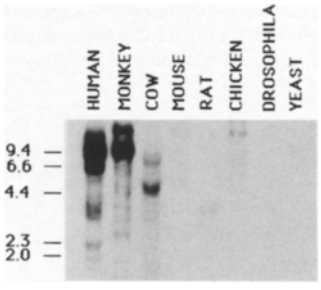

Fig. 2. Evolutionary conservation of C2H2-25 gene, Southern blot analysis. DNA isolated from the indicated species (Clontech, Palo Alto, Calif.) was digested with *Eco*RI. The DNA (10 μg/lane) was run on a 1% agarose gel in 1× TBE, and transferred to a nylon filter (MSI, Westboro, Mass.). C2H2-25 cDNA insert was radiolabeled with ^{32}P by random priming (Prime-It, Stratagene) and used to probe Southern blots in QuikHyb hybridization solution (Stratagene) at 68°C. Blots were washed twice at room temperature with 2× SSC, 0.1% SDS, followed by two washes at 60°C with 1× SSC, 0.1% SDS, and autoradiography was performed. Molecular weight markers (bp) are indicated on the left.

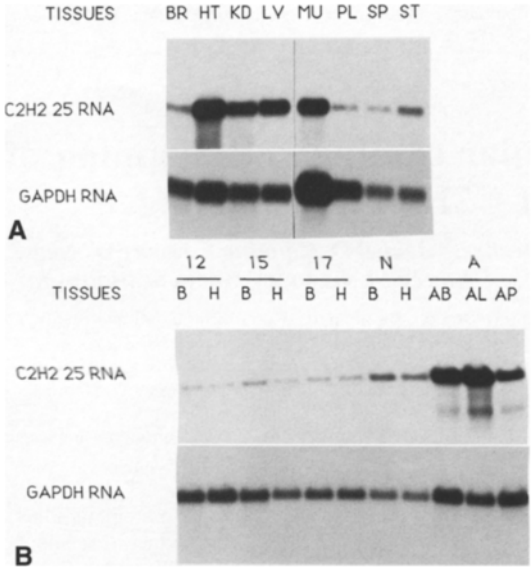

Fig. 3. RNA expression of C2H2-25, Northern blot analysis. **(A)** Tissue expression of C2H2-25 RNA. Poly $(A)^+$ mRNA (4 μg) from various adult mouse tissues (BR, brain; HT, heart; KD, kidney; LV, liver; MU, muscle; PL, placenta; SP, spleen; ST, stomach) was probed with a fragment containing the complete coding sequence of the C2H2-25 cDNA. **(B)** Developmental expression of C2H2-25 mRNA. Poly $(A)^+$ mRNA from different embryonic mouse developmental stages was probed with the complete coding sequence for the C2H2-25 cDNA. Body and head samples are shown. The gestational age is given in days. Adult liver (AL), adult placenta (AP), and adult brain (AB) are also shown. In both Frames A and B, GAPDH cDNA was used to probe Northern blots as a control for the amount of RNA loaded in individual lanes and the integrity of the RNA.

C2H2-25 gene exists as a single copy in the human genome (L. Stubbs, personal communication). Northern blot analysis indicated that C2H2-25 is ubiquitously expressed in all adult mouse tissues examined (Fig. 3A). In addition, C2H2-25 is expressed early during mouse development, increases in expression during development, and is expressed at maximal levels in adult tissues (Fig. 3B).

The gene encoding C2H2-25 was mapped to Chr 19 by PCR analysis of the NIGMS human-rodent somatic cell hybrid panel #1 and the monochromosomal NIGMS human-rodent somatic cell hybrid panel #2 (NIGMS, Camden, N.J.) (Polymeropoulos et al. 1991). The map location on Chr 19 was sub-localized by PCR amplification of a contigued Mega Yac library (Research Genetics, Huntsville, Ala.) (Berry et al. 1995). C2H2-25 is contained on the YAC 965_C_8. This YAC also contains the marker D19S218, whose cytogenetic location has been determined to be 19q13.4. We have used the following primers located in the 3′ UTR of C2H2-25 in the PCR analyses described above to map the location of the C2H2-25 gene: 5′-TGTCACTGACAGTTTCTGAGGCAG-3′; 5′-GTCAGAGAGGAAGACTCAGACTAT-3′. In support of our mapping data, others have independently mapped C2H2-25 to human Chr 19q13.4 by somatic cell hybrid analysis with primers from the 3′ UTR of C2H2-25 (L. Stubbs, personal communication). Many C2H2-type zinc fingers have previously been shown to be important in developmental processes. A number of developmental anomalies have been mapped to 19q including autosomal nonsyndromic sensorineural deafness (19q; Chen et al. 1995), orofacial cleft-3 (19q13; Stein et al. 1995), cone-rod retinal dystrophy (19q13.1-q13.2; Evans et al. 1994), and retinitis pigmentosa-II (19q13.4; Al-Maghtheh et al. 1994). Thus, aberrant expression of C2H2-25 may contribute to these developmental anomalies.

References

Al-Maghtheh M, Ingelhearn CF, Keen TJ, Evans K, Moore AT, Jay M, Bird AC, Bhattacharya SS (1994) Identification of a sixth locus for

autosomal dominant retinitis pigmentosa on chromosome 19. Hum Mol Genet 3, 351–354
Beato M (1989) Gene regulation by steroid hormones. Cell 56, 335–344
Becker KG, Nagle JW, Canning RD, Biddison WE, Ozato K, Drew PD (1995) Rapid isolation and characterization of 118 novel C2H2-type zinc finger cDNAs expressed in human brain. Hum Mol Genet 4, 685–691
Bellefroid EJ, Lecocq PJ, Benhida A, Poncelet DA, Belayew A, Martial JA (1989) The human genome contains hundreds of genes coding for zinc finger proteins of the Kruppel type. DNA 8, 377–387
Bellefroid EJ, Poncelet DA, Lecocq PJ, Revelant O, Martial JA (1991) The evolutionarily conserved Kruppel-associated box domain defines a subfamily of eukaryotic multifingered proteins. Proc Natl Acad Sci USA 88, 3608–3612
Berry R, Walter NAR, Stevens TJ, Rubano T, Wilcox AS, Hopkins JA, Sikela JM (1995) A strategy for the identification of candidate genes for alcohol related phenotypes and other human disorders using rapid polymerase chain reaction mapping of gene based sequence tagged sites. Alcohol Clin Exp Res 19, 821–823
Brown RS, Sander C, Argos P (1985) The primary structure of transcription factor TFIIIA has 12 consecutive repeats. FEBS Lett 186, 271–274
Chen AH, Ni L, Fukushima K, Marietta J, O'Neill M, Coucke P, Willems P, Smith RJH (1995) Linkage of a gene for dominant non-syndromic deafness to chromosome 19. Hum Mol Genet 4, 1073–1076
Evans K, Fryer A, Inglehearn C, Duvall-Young J, Whittaker JL, Gregory CY, Butler R, Ebenezer N, Hunt DM, Bhattacharya S (1994) Genetic linkage of cone-rod retinal dystrophy to chromosome 19q and evidence for segregation distortion. Nature Genet 6, 210–213
Fortini ME, Lai Z, Ruben GE (1991) The Drosophila znf-1 and znf-2 genes encode novel proteins containing both zinc-finger and homeodomain motifs. Mech Dev 34, 113–122
Freemont PS (1993) The ring finger: a novel protein sequence motif related to the zinc finger. Ann NY Acad Sci 684, 174–192
Ginsberg AM, King BO, Roeder RG (1984) Xenopus 5S gene transcription factor, TFIIIA: characterization of a cDNA clone and measurement of RNA levels throughout development. Cell 39, 479–489
Harrison S, Travers A (1990) The tramtrack gene encodes a Drosophila finger protein that interacts with the ftz transcriptional regulatory region and shows a novel embryonic expression pattern. EMBO J 9, 207–216
Lee J-S, Galvin KM, Shi Y (1993) Evidence for physical interaction between the zinc-finger transcription factors YY1 and SP1. Proc Natl Acad Sci USA 90, 6145–6149
Miller J, McLachlan AD, Klug A (1985) Repetitive zinc-binding domains in the protein transcription factor TFIIIA from *Xenopus laevis*. EMBO J 4, 1609–1614
Pieler T, Bellefroid E (1994) Perspectives on zinc finger protein function and evolution—an update. Mol Biol Rep 20, 1–8
Polymeropoulos MH, Xiao H, Glodek M, Adams MD, Moreno RF, Fitzgerald MG, Venter JC, Merril CR (1991) Chromosomal assignment of 46 brain cDNAs. Genomics 12, 492–49
Schuh R, Aicher W, Gaul U, Cote S, Preiss A, Maier D, Seifert E, Nauber U, Schroder C, Kemler R, Jackle H (1986) A conserved family of nuclear proteins containing structural elements of finger protein encoded by Kruppel, a Drosophila segmentation gene. Cell 47, 1025–1032
Stein J, Mulliken JB, Stal S, Gasser DL, Malcolm S, Winter R, Blanton SH, Amos C, Seemanova E, Hecht JT (1995) Nonsyndromic cleft lip with or without cleft palate: evidence of linkage to BCL3 in 17 multigenerational families. Am J Hum Genet 57, 257–272
Tommerup N, Vissing H (1995) Isolation and fine mapping of 16 novel human zinc finger-encoding cDNAs identify putative candidate genes for developmental and malignant disorders. Genomics 27, 259–264